# A BRIEF INTRODUCTION TO BANACH SPACES, THEIR DUALS, AND $L^p$ SPACES: GRADUATE LECTURE COURSE, AUTUMN TERM 2001, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF YORK

S. P. EVESON

## NOTATION

$\mathbb{R}$ and $\mathbb{C}$ mean the real and complex fields, as usual. We use the symbol $\mathbb{K}$ to mean 'either $\mathbb{R}$ or $\mathbb{C}$'. Sequences will normally be denoted using subscripts, as in $(x_n)$, but where this clashes with the conventional notation for the components of a vector in $\mathbb{K}^N$, a bracketed superscript will be used instead, so $x_j^{(n)}$ means the $j$th component of the $n$th vector in the sequence. The same notation will be used for a sequence of sequences, so $x^{(n)}$ will be the $n$th sequence and $x_j^{(n)}$ the $j$th term of the $n$th sequence.

Using this notation, the $n$th standard basis vector $e^{(n)}$ is defined by

$$e_j^{(n)} = \begin{cases} 1 & \text{if } j = n \\ 0 & \text{if } j \neq n. \end{cases}$$

We write $e^{(n)}$ for either a vector in $\mathbb{K}^N$ or an element of a sequence space; it will be clear from the context which is meant.

If $(X, d)$ is a metric space, $x \in X$ and $Y \subseteq X$ we define the distance of $x$ from $Y$ by

$$\operatorname{dist}(x, Y) = \inf\{d(x, y) : y \in Y\}.$$

Note that $\operatorname{dist}(x, Y) = 0$ if and only if $x \in \overline{Y}$ (the closure of $Y$).

The *characteristic function* of a set $A$ is defined on any superset of $A$ by

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

Recall that a metric space is called *separable* if it contains a countable dense subset.

## 1. NORMED AND BANACH SPACES

1.1. **Definition.** Let $X$ be a vector space over $\mathbb{K}$. A *norm* on $X$ is a function $\|\cdot\| : X \to \mathbb{R}$ such that

(1) if $x, y \in X$ then $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality);
(2) if $x \in X$ and $\lambda \in \mathbb{K}$ then $\|\lambda x\| = |\lambda| \|x\|$;
(3) if $x \in X$ and $\|x\| = 0$ then $x = 0$.

A vector space equipped with a norm is called a *normed space*. Associated with the norm is a metric $d$ defined by $d(x, y) = \|x - y\|$.

1.2**. Exercise.** Verify that $\|x\| \geq 0$ for all $x \in X$, that $\|0\| = 0$, and that $d$ is indeed a metric. Prove the useful *reverse triangle inequality*:

$$\big|\|x\| - \|y\|\big| \leq \|x \pm y\|.$$

Unless otherwise stated, all ideas of convergence, continuity, completeness, compactness, etc. in $X$ are defined in terms of this metric. For example, a sequence $(x_n)$ is Cauchy in $X$ if for any $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that if $m, n > N$ then $\|x_m - x_n\| < \varepsilon$, and a sequence $(x_n)$ converges to a limit $x$ in $X$ if for any $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that if $n > N$ then $\|x_n - x\| < \varepsilon$. Infinite series in normed spaces are defined in exactly the same way as in $\mathbb{R}$ or $\mathbb{C}$: we write $\sum_{k=1}^{\infty} x_k = x$ to mean that $\sum_{k=1}^{n} x_k \to x$ as $n \to \infty$.

1.3**. Definition.** A normed space $X$ is called a *Banach space* (abbr. B-space) if it is complete, i.e. if every Cauchy sequence in $X$ converges to a limit in $X$.

1.4**. Theorem.** *A normed space $X$ is complete (i.e. a B-space) if and only if every absolutely convergent series converges; precisely, if the convergence of the series $\sum_{k=1}^{\infty} \|x_k\|$ implies the convergence of the series $\sum_{k=1}^{\infty} x_k$.*

*Proof.* The first half is identical to the usual proof for $\mathbb{R}$ or $\mathbb{C}$. Suppose that $X$ is complete and that $\sum_{k=1}^{\infty} \|x_k\|$ converges. Then the partial sums form a Cauchy sequence, so for any $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that if $m, n > N$ then $\sum_{k=m}^{n} \|x_k\| < \varepsilon$. But

$$\left\| \sum_{k=m}^{n} x_k \right\| \leq \sum_{k=m}^{n} \|x_k\| < \varepsilon$$

by the triangle inequality, so the partial sums of $\sum_{k=1}^{\infty} x_k$ also form a Cauchy sequence, which converges since $X$ is complete.

For the other half, suppose that all absolutely convergent series in $X$ converge and that $(x_n)$ is a Cauchy sequence. We can choose $n_0$ such that if $n > n_0$ then $\|x_n - x_{n_0}\| < 1/2$. Similarly, we can choose $n_1 > n_0$ such that if $n > n_1$ then $\|x_n - x_{n_1}\| < 1/4$, and so on to give a sequence $(n_k)$ with the property that $\|x_{n_k} - x_{n_{k-1}}\| < 1/2^k$. In particular, the series

$$\sum_{k=1}^{\infty} x_{n_k} - x_{n_{k-1}}$$

is absolutely convergent, hence convergent by hypothesis. But the $k$th partial sum of this series is $x_{n_k} - x_{n_0}$, so $x_{n_k}$ converges and $(x_n)$ has a convergent subsequence. But (**exercise**) if a Cauchy sequence has a convergent subsequence then it is itself convergent, so $(x_n)$ is convergent and $X$ is complete. □

1.5**. Exercise.** A series $\sum_{k=1}^{\infty} x_k$ in a B-space is called *unconditionally convergent* if for every permutation $\pi$ of $\mathbb{N}$ (i.e., every bijection from $\mathbb{N}$ to itself),

$$\sum_{k=1}^{\infty} x_{\pi(k)} = \sum_{k=1}^{\infty} x_k.$$

Show that every absolutely convergent series in a B-space is unconditionally convergent (unlike in the scalar case, the converse is not true: see Exercise 2.12).

1.6. **Definition.** Two norms $\|\cdot\|$ and $\|\cdot\|'$ on a vector space $X$ are called *equivalent* if there exist constants $\alpha, \beta > 0$ such that for all $x \in X$, $\alpha\|x\|' \leq \|x\| \leq \beta\|x\|'$.

As its name suggests, equivalence of norms is an equivalence relation. Equivalent norms share the same Cauchy and convergent sequences, continuous functions, compact sets, etc. In particular, if $X$ is complete in a given norm, then it is complete in all equivalent norms.

1.7. **Theorem.** *If $X$ is a finite-dimensional vector space over $\mathbb{K}$, then all norms on $X$ are equivalent. In particular, any norm on $X$ makes it into a B-space.*

*Proof.* Any $N$-dimensional vector space over $\mathbb{K}$ is isomorphic to $\mathbb{K}^N$, and $\mathbb{K}^N$ is complete in the Euclidean norm. It is therefore sufficient to show that an arbitrary norm $\|\cdot\|$ on $\mathbb{K}^N$ is equivalent to the Euclidean norm, $\|\cdot\|_2$. For any $x \in \mathbb{K}^N$, $x = \sum_{j=1}^N x_j e^{(j)}$. By the triangle and Cauchy-Schwarz inequalities,

$$\|x\| \leq \sum_{j=1}^N |x_j|\|e_j\| \leq \|x\|_2 \left(\sum_{j=1}^N \|e^{(j)}\|^2\right)^{1/2}$$

which is one of the required inequalities. This estimate, together with the reverse triangle inequality, also shows that $\|\cdot\|$ is a continuous function in the Euclidean metric. By the Heine-Borel Theorem, $\|\cdot\|$ attains its infimum $\alpha$ on the sphere $\{x \in \mathbb{K}^N : \|x\|_2 = 1\}$. This minimum cannot be zero, otherwise we would have $\|x\| = 0$ for some non-zero $x$. Now, for any $x \neq 0$ we have $0 < \alpha \leq \|x/\|x\|_2\|$, or $\alpha\|x\|_2 \leq \|x\|$ which is the other required inequality. $\square$

## 2. Examples of Normed and Banach Spaces

2.1. **Example.** Inner products naturally give rise to norms. Recall that if $X$ is a vector space over $\mathbb{K}$, then a function $\langle\cdot,\cdot\rangle : X \times X \to \mathbb{K}$ is called an *inner product* if for all $x, y, z \in X$ and $\lambda, \mu \in \mathbb{K}$ we have

(1) $\langle \lambda x + \mu y, z\rangle = \lambda \langle x, z\rangle + \mu \langle y, z\rangle$;
(2) $\langle x, y\rangle = \overline{\langle y, x\rangle}$;
(3) $\langle x, x\rangle \geq 0$ and if $\langle x, x\rangle = 0$ then $x = 0$.

The norm associated with the inner product is given by $\|x\| = \langle x, x\rangle^{1/2} \geq 0$. A complete inner product space is a particular kind of B-space called a *Hilbert space*.

2.2. **Exercise.** Prove the Cauchy-Schwarz inequality: if $x$ and $y$ are vectors in an inner product space then $|\langle x, y\rangle| \leq \|x\|\|y\|$. Use this to show that $\|\cdot\|$ defined above really is a norm.

*Suggestion: consider $\langle x + \lambda y, x + \lambda y\rangle$ for real $\lambda$. This is a quadratic with at most one root, which implies an inequality between its coefficients which turns out to be almost, but not quite, the Cauchy-Schwarz inequality. To complete the argument, multiply $x$ or $y$ by a suitable scalar of modulus* 1.

2.3. **Example.** The finite-dimensional vector space $\mathbb{K}^N$ can be turned into a B-space in many different ways. For example, define

$$\begin{aligned}\|x\|_p &= \left(\sum_{j=1}^N |x_j|^p\right)^{1/p} \qquad (p \in [1,\infty)) \\ \|x\|_\infty &= \max_{j=1}^N |x_j|.\end{aligned}$$

If $p = 2$, this is the ordinary Euclidean norm. If $p = 1$ or $p = \infty$ it is easy to check directly that $\|\cdot\|_p$ is a norm. For other values of $p$ everything is easy except the triangle inequality, which is Minkowski's inequality (Corollary 2.6). All these norms are complete by Theorem 1.7.

2.4. **Exercise.** Show that if $x \in \mathbb{R}^N$ then $\|x\|_p$ is a decreasing function of $p$ which tends to $\|x\|_\infty$ as $p \to \infty$ (hence the notation). *Suggestion: when you compare* $\|x\|_{p_1}$ *with* $\|x\|_{p_2}$*, it can make life easier if you scale* $x$ *so one of the norms is* $1$.

It remains to establish the triangle inequality for $\|\cdot\|_p$, for which we need another important inequality.

2.5. **Theorem** (Finite-Dimensional Hölder Inequality). *Fix* $p \in (1,\infty)$ *and let* $q = p/(p-1)$*, so* $1/p + 1/q = 1$ *(*$q$ *is the* (Hölder) conjugate *of* $p$*). Then for any* $x, y \in \mathbb{K}^N$ *we have*

$$\left|\sum_{j=1}^N x_j y_j\right| \le \sum_{j=1}^N |x_j y_j| \le \|x\|_p \|y\|_q.$$

*(Note that if* $p = 2$ *then* $q = 2$ *and this is a special case of the Cauchy-Schwarz inequality).*

*Proof.* The first inequality is just the scalar triangle inequality; the second is the interesting one. It is trivial if $x = 0$ or $y = 0$, so assume not. We begin with an auxiliary inequality. Since $\exp$ is a convex function and $1/p + 1/q = 1$, we have for any $\alpha, \beta \in \mathbb{R}$

$$\frac{1}{p}\exp(\alpha) + \frac{1}{q}\exp(\beta) \ge \exp\left(\frac{1}{p}\alpha + \frac{1}{q}\beta\right) = \exp(\alpha/p)\exp(\beta/q)$$

with equality if and only if $\alpha = \beta$. Now, if $a$ and $b$ are positive real numbers, substitute $\alpha = p\log a$ and $\beta = q \log b$ to give

$$ab \le \frac{a^p}{p} + \frac{b^q}{q}. \tag{*}$$

Moreover (*) is (trivially) true if $a = 0$ or $b = 0$. Now, let $\xi_j = |x_j|/\|x\|_p$ and $\eta_j = |y_j|/\|y\|_q$, so $\xi_j \ge 0$, $\sum_{j=1}^N \xi_j^p = 1$, $\eta_j \ge 0$ and $\sum_{j=1}^N \eta_j^q = 1$. Applying (*) to $\xi_j$ and $\eta_j$ and summing gives us

$$\sum_{j=1}^N \xi_j \eta_j \le \frac{1}{p}\sum_{j=1}^N \xi_j^p + \frac{1}{q}\sum_{j=1}^n \eta_j^q = \frac{1}{p} + \frac{1}{q} = 1.$$

Finally, substituting the definitions of $\xi_j$ and $\eta_j$ gives

$$\sum_{j=1}^N |x_j||y_j| \le \|x\|_p \|y\|_q.$$

□

2.6. **Corollary** (Finite-Dimensional Minkowski Inequality). *Suppose $x, y \in \mathbb{R}^N$ and $p \in (1, \infty)$. Then*

$$\|x + y\|_p \leq \|x\|_p + \|y\|_p.$$

*Proof.* This is trivial if $x + y = 0$, so assume not. We have

$$\sum_{j=1}^{N} |x_j + y_j|^p = \sum_{j=1}^{N} |x_j + y_j||x_j + y_j|^{p-1} \leq \sum_{j=1}^{N} |x_j||x_j + y_j|^{p-1} + \sum_{j=1}^{N} |y_j||x_j + y_j|^{p-1}.$$

Applying Hölder's inequality to both sums on the RHS and collecting terms gives

$$\|x + y\|_p^p \leq (\|x\|_p + \|y\|_p) \left( \sum_{j=1}^{N} |x_j + y_j|^{(p-1)q} \right)^{1/q}.$$

Now, $(p-1)q = p$, so the sum on the RHS is in fact $\|x+y\|_p^{p/q}$. Finally, $p - p/q = 1$, so dividing through by this (non-zero by hypothesis) term gives

$$\|x + y\|_p \leq \|x\|_p + \|y\|_p.$$

□

2.7. **Exercise.** Under what circumstances do we have equality in either of these inequalities?

Note the symmetry in the above definition: if $q$ is conjugate to $p$ then $p$ is conjugate to $q$. If $p < 2$ then $q > 2$ and as $p$ approaches 1, $q$ approaches $\infty$; we often extend the idea of conjugacy and say that 1 and $\infty$ are conjugate to each other, though many facts that are true in $(1, \infty)$ break down to a greater or lesser extent at the endpoints.

2.8. **Exercise.** Show that the Hölder and Minkowski inequalities remain true if $p = 1$ and $q = \infty$ (or vice-versa).

Now for some infinite-dimensional examples.

2.9. **Example.** If $p \in [1, \infty)$, we define the space $\ell^p$ ($\ell^p_{\mathbb{R}}$ or $\ell^p_{\mathbb{C}}$ if we want to emphasise the field) to be the space of all sequences $(x_j)$ in $\mathbb{K}$ such that $\sum_{j=1}^{\infty} |x_j|^p$ converges. The norm is defined by

$$\|x\|_p = \left( \sum_{j=1}^{\infty} |x_j|^p \right)^{1/p}.$$

The space $\ell^\infty$ is the space of all bounded sequences, with norm

$$\|x\|_\infty = \sup_{j \in \mathbb{N}} |x_j|.$$

We should check that these really are vector spaces, and that the formulae above really do define norms. For example, to show that $\ell^p$ is closed under addition if $p \in [1, \infty)$ we proceed as follows:

if $x, y \in \ell^p$, then for any $n$,

$$\left(\sum_{j=1}^{n} |x_j + y_j|^p\right)^{1/p} \leq \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p} + \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p}$$
$$\leq \|x\|_p + \|y\|_p$$

so

$$\sum_{j=1}^{n} |x_j + y_j|^p \leq (\|x\|_p + \|y\|_p)^p$$

and by the principle of bounded monotone convergence, $\sum_{j=1}^{\infty} |x_j + y_j|^p$ converges and

$$\sum_{j=1}^{\infty} |x_j + y_j|^p \leq (\|x\|_p + \|y\|_p)^p.$$

This shows that $x + y \in \ell^p$ and that $\|x + y\|_p \leq \|x\|_p + \|y\|_p$, so the triangle inequality (infinite-dimensional Minkowksi inequality) comes for free. Other details left as an **exercise** (including the $p = \infty$ case).

2.10**. Exercise.** (cf. Exercise 2.4). Show that:

(1) if $1 \leq p_1 \leq p_2$ and $x \in \ell^{p_1}$ then $x \in \ell^{p_2}$ and $\|x\|_{p_2} \leq \|x\|_{p_1}$;
(2) if $x \in \ell^{p_0}$ for some $p_0$ then $x \in \ell^\infty$ and $\|x\|_p \to \|x\|_\infty$ as $p \to \infty$;
(3) no two $\ell^p$ spaces coincide.

2.11**. Exercise.** Establish the infinite-dimensional Hölder inequality: if $p$ and $q$ are conjugate, $x \in \ell^p$ and $y \in \ell^q$ then

$$\left|\sum_{j=1}^{\infty} x_j y_j\right| \leq \|x\|_p \|y\|_q.$$

In the example above, we used the finite-dimensional Minkowski inequality to prove the infinite-dimensional one. Give an alternative proof based on the infinite-dimensional Hölder inequality.

2.12**. Exercise.** (cf. Exercise 1.5). Show that if $p \in (1, \infty)$ then $\sum_{n=1}^{\infty} e^{(n)}/n$ is unconditionally convergent but not absolutely convergent.

2.13**. Exercise.** We shall see later a general result (Theorem 3.10) which includes the fact that $\ell^p$ is complete as a special case. Give a direct proof of this result.

*Suggestion: if $1 \leq p < \infty$, take a Cauchy sequence $(x^{(n)})$ in $\ell^p$ and show that for each $j$, $(x_j^{(n)})$ is Cauchy in $j$, hence convergent to $x_j$, say. In the Cauchy condition for $(x^{(n)})$, replace the infinite sum in $\|x^{(m)} - x^{(n)}\|_p$ by a finite one, and send things off to infinity in just the right order. The $p = \infty$ case is similar but easier.*

2.14**. Example.** Let $S$ be a (Lebesgue) measurable subset of $\mathbb{R}^N$ and $p \in [1, \infty)$, and define the space $\mathscr{L}^p(S)$ ($\mathscr{L}^p_{\mathbb{R}}(S)$ or $\mathscr{L}^p_{\mathbb{C}}(S)$ if you want to be precise about the field) of measurable functions $f : S \to \mathbb{K}$ such that $\int_S |f|^p$ is finite. We can try to define a norm by

$$\|f\|_p = \left( \int_S |f|^p \right)^{1/p}$$

but this does not quite work. We can easily generalise Hölder's inequality: if $p$ and $q$ are conjugate, $f \in \mathscr{L}^p(S)$ and $g \in \mathscr{L}^q(S)$ then $fg$ is integrable over $S$ and

$$\int_S fg \leq \|f\|_p \|g\|_q$$

and Minkowski's inequality: $\mathscr{L}^p(S)$ is indeed a vector space, and $\|\cdot\|_p$ satisfies the triangle inequality. However, there is one property missing: if $\|f\|_p = 0$, then $\int_S |f|^p = 0$, so $|f|^p = 0$ almost everywhere and hence $f = 0$ a.e. But this certainly does not imply that $f$ is identically zero. To fix this problem, we introduce the subspace $\mathscr{N}$ of *null functions* (functions which are almost everywhere zero), and consider the quotient space $\mathscr{L}^p(S)/\mathscr{N}$. We define the norm of a coset by $\|f + \mathscr{N}\|_p = \|f\|_p$, and call this new space $L^p(S)$.

2.15**. Exercise.** Supply the details in the construction of $L^p(S)$.

2.16**. Exercise.** Suppose $p, q \in (1, \infty)$ are such that for all $f \in L^p(\mathbb{R})$ and $g \in L^q(\mathbb{R})$, $\int_{\mathbb{R}} fg \leq \|f\|_p \|g\|_q$. Show that $1/p + 1/q = 1$. Is the same true for $\ell^p$?

2.17**. Exercise.** Suppose that $S \subseteq \mathbb{R}^N$ has finite measure $M$. Use Hölder's inequality to show that if $f \in L^p(S)$ for some $p \in (1, \infty)$ then $f \in L^1(S)$ and

$$\|f\|_1 \leq M^{1/q} \|f\|_p$$

where $q$ is conjugate to $p$. Now suppose that $1 < p_1 < p_2 < \infty$ and that $g \in L^{p_2}(S)$. Substitute $f = |g|^{p_2}$ and $p = p_2/p_1$ into this estimate to show that $g \in L^{p_1}(S)$, and obtain an estimate for $\|f\|_{p_1}$. What about $L^\infty(S)$?

Note that the inclusions for $L^p(S)$ run the *opposite way* to those for $\ell^p$ (Exercise 2.10) when $S$ has finite measure. When $S$ has infinite measure, it's different again: give examples to show that if $p_1 \neq p_2$ then neither of the spaces $L^{p_1}(\mathbb{R})$, $L^{p_2}(\mathbb{R})$ is included in the other.

2.18**. Example.** There is also an $L^\infty$ space to go with the other $L^p$ spaces. If $f$ is a real-valued function on a measurable subset $S$ of $\mathbb{R}^N$, we say that $f$ is *essentially bounded above* on $S$ if there is a constant $C$ (an *essential upper bound*) such that $f(s) \leq C$ for almost all $s \in S$. The *essential supremum* $\operatorname{ess\,sup}_S f$ of such a function is the smallest essential upper bound. Essential lower bounds and the essential infimum are defined in a similar way.

Now, we define $\mathscr{L}^\infty(S)$ to be the space of all essentially bounded, measurable functions on $S$, with norm

$$\|f\|_\infty = \operatorname*{ess\,sup}_S |f|.$$

As before, this is not a norm because of the existence of null functions, but $L^\infty(S) = \mathscr{L}^\infty(S)/\mathscr{N}$ is a B-space with $\|f + \mathscr{N}\|_\infty = \|f\|_\infty$.

2.19**. Exercise.** Fill in the details in the construction of $L^\infty$ and the proof of its completeness.

2.20**. Exercise.** Show that the following are all B-spaces when endowed with the $\infty$ norm (the supremum of the absolute value of the function or sequence).

(1) the space of continuous $\mathbb{K}$-valued functions on a compact metric space (or indeed a compact Hausdorff space);
(2) the space $c$ of all convergent sequences in $\mathbb{K}$;
(3) the space $c_0$ of all sequences converging to 0 in $\mathbb{K}$;
(4) the space of all bounded $\mathbb{K}$-valued functions on an arbitrary set (generalising $\ell^\infty$).

## 3. An aside on measure theory

Measure theory is an abstraction of the ideas of length, area, volume and cardinality. This is not a course about measure theory, but we need to know a little about it. This section more or less follows Chapter 1 of Rudin [5], to which you should refer for further details (and most of the proofs!)

Before we begin, we introduce some conventions about $\infty$. We consider the set $[0, \infty]$, which contains all the non-negative reals and a symbol denoted by $\infty$.

We then make $\infty$ obey a few natural rules: $x + \infty = \infty$ and $x \times \infty = \infty$ for all $x \in (0, \infty]$, $0 \times \infty = 0$ and $x < \infty$ for all $x \in [0, \infty)$. As long as we do not perform any other operations on $\infty$, this is all perfectly consistent: (**exercise**) the associative, commutative and distributive laws remain true. We must not, however, do anything else: in particular, $\infty - \infty$ and $\infty/\infty$ are illegal combinations of symbols.

One consequence of these conventions is that identities like

$$\sum_{n=1}^{\infty}(a_n + b_n) = \sum_{n=1}^{\infty} a_n + \sum_{n=1}^{\infty} b_n$$

work even for divergent series (of non-negative terms).

3.1**. Definition.** A *measure space* is conventionally (see Rudin's 'comment on terminology', §1.21) described as consisting of three pieces. Firstly, we have a set $S$. Secondly, we have a collection $\Sigma$ of subsets of $S$ with the following properties:

(1) $S \in \Sigma$;
(2) if $A \in \Sigma$ then $S \setminus A \in \Sigma$;
(3) if $(A_n)$ is a sequence in $\Sigma$ then $\bigcup_{n\in\mathbb{N}} A_n \in \Sigma$

(from which it immediately follows that $\emptyset \in \Sigma$ and that $\Sigma$ is closed under countable intersections). Such a collection is called a *σ-field* or *σ-algebra* on $S$. Finally, we have a function $\mu$ from $\Sigma$ to $[0, \infty]$ such that $\mu(\emptyset) = 0$ and if $(A_n)$ is a sequence of pairwise disjoint sets in $\Sigma$, then

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

(*countable additivity* or *σ-additivity*), where the sum of the series is to be interpreted as $\infty$ if it diverges.

The triple $(S, \Sigma, \mu)$ is called a *measure space*, elements of $\Sigma$ are called *measurable sets* and $\mu$ is a *positive measure*. The measure (or the measure space) is called *finite* or *bounded* if $\mu(S) < \infty$ and *$\sigma$-finite* if there is a sequence $(A_n)$ of sets in $\Sigma$ with $\mu(A_n) < \infty$ for each $n$ and $S = \bigcup_{n\in\mathbb{N}} A_n$.

A set $N \in \Sigma$ is called *$\mu$-null* if $\mu(N) = 0$. A property that is true for all elements of $S$ except those in a $\mu$-null set is said to hold *$\mu$-almost everywhere* or *$\mu$-a.e.*

3.2**. Example.** The simplest possible measure is called a *Dirac measure* or *point mass*. We take any set $S$, any $\sigma$-field $\Sigma$ on $S$ and any point $x \in S$. We then define

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

It is easy to check (**exercise**) that this is a finite measure. The next simplest example is *counting measure*. Again take any set $S$ and any $\sigma$-field $\Sigma$ and define $\mu(A)$ to be the number of elements in $A$, or $\infty$ if $A$ is infinite. Again it it easy to check (**exercise**) that this is a measure which is finite if and only if $S$ is finite and $\sigma$-finite if and only if $S$ is countable.

More complicated examples require quite a lot more work. On $\mathbb{R}$, let $\Sigma$ be the smallest $\sigma$-field containing all open intervals $(a, b)$ (elements of this field are called *Borel sets*). It can be shown, with a certain amount of work, that there is a unique positive measure $\mu$ defined on $\Sigma$ such that $\mu(a, b) = b-a$. This is (1-dimensional) *Lebesgue measure*. It is not finite but is $\sigma$-finite (consider the intervals $[n, n+1)$ for $n \in \mathbb{Z}$). Similarly, in $\mathbb{R}^2$ we can take the $\sigma$-field generated by open rectangles and it turns out that there is a unique measure on this field such that the measure of each rectangle is its area; this is 2-dimensional Lebesgue measure. All this can be done in $\mathbb{R}^N$ for any $N \in \mathbb{N}$, but the contruction is outside the scope of this course.

We can also add and scale measures in the natural way: if $\mu$ and $\nu$ are positive measures on the same $\sigma$-field and $c > 0$, we can define $(\mu + \nu)(A) = \mu(A) + \nu(A)$ and $(c\mu)(a) = c\mu(A)$ to obtain new positive measures. Subtraction requires a slightly different setup: see Definition 3.12.

Once we have a positive measure, we can construct an integral. We begin with a small class of functions, then extend.

3.3**. Definition.** Let $(S, \Sigma, \mu)$ be a measure space. A function $g : S \to \mathbb{K}$ is called *simple* if it can be written in the form

$$g = \sum_{j=1}^{n} c_j \chi_{A_j}$$

where $A_j \in \Sigma$ for all $j$. The integral of such a function is defined by

$$\int_S g \,\mathrm{d}\mu = \sum_{j=1}^{n} c_j \mu(A_j)$$

(remember that $0 \times \infty = 0$). A function $f : S \to \mathbb{K}$ is called *measurable* if $f^{-1}(A) \in \Sigma$ whenever $A$ is an open subset of $\mathbb{K}$.

Next, the integral of a non-negative measurable function $f$ is defined by

$$\int_S f \,\mathrm{d}\mu = \sup_{\substack{0 \le g \le f \\ g \text{ simple}}} \int_S g \,\mathrm{d}\mu .$$

Finally, we define the integral of a measurable, complex-valued function. Write

$$h^+(t) = \begin{cases} h(t) & \text{if } h(t) \ge 0 \\ 0 & \text{if } h(t) < 0 \end{cases} \qquad h^-(t) = \begin{cases} 0 & \text{if } h(t) \ge 0 \\ -h(t) & \text{if } h(t) < 0 \end{cases}$$

and define

$$\int_S f \,\mathrm{d}\mu = \int_S (\operatorname{Re} f)^+ \,\mathrm{d}\mu - \int_S (\operatorname{Re} f)^- \,\mathrm{d}\mu + \mathrm{i} \int_S (\operatorname{Im} f)^+ \,\mathrm{d}\mu - \mathrm{i} \int_S (\operatorname{Im} f)^- \,\mathrm{d}\mu$$

provided all of these integrals are finite.

Integration defined in this way has all the expected properties: it is linear; the values of a function on a $\mu$-null set do not affect its integrability or the value of its integral; it is positive (i.e. if $f \le g$ $\mu$-a.e. then $\int_S f \,\mathrm{d}\mu \le \int_S g \,\mathrm{d}\mu$); and if $f \ge 0$ $\mu$-a.e. and $\int_S f \,\mathrm{d}\mu = 0$ then $f = 0$ $\mu$-a.e. Moreover, if $f$ is measurable, $g$ is integrable and $|f| \le g$ $\mu$-a.e. then $f$ is integrable.

3.4**. Example.** Integration with respect to Lebesgue measure is, of course, Lebesgue integration.

Consider a $\sigma$-algebra $\Sigma$ on a set $S$, a Dirac measure $\delta_x$ such that the singleton $\{x\}$ is measurable, and a non-negative measurable function $f$. If $g$ is any simple function with $g \le f$, then we can write

$$g = \sum_{j=1}^{n} c_j \chi_{A_j}$$

where the $(A_j)$ are disjoint and measurable. We claim that $\int_S g \le f(x)$. This is trivial if $x$ does not lie in any of the $(A_j)$, for then $\int_S g = 0$. If, on the other hand, $x \in A_j$ for some $j$ then $\int_S g = c_j$ and $c_j = g(x) \le f(x)$. This shows that $\int f \le f(x)$. Moreover, $f(x)\chi_{\{x\}} \le f$ and $\int_S f(x)\chi_{\{x\}} = f(x)$, so $\int_S f \ge f(x)$. Combining these, we see that $\int_S f = f(x)$. The integral is thus determined by the value of $f$ at one point, in strong contrast to the Lebesgue integral. It is easy to check (**exercise**) that the same is true for complex-valued functions.

Now consider the natural numbers, the $\sigma$-field of all subsets of $\mathbb{N}$, and counting measure $\mu$. We claim that if $f$ is a non-negative function on $\mathbb{N}$ then

$$\int f \,\mathrm{d}\mu = \sum_{n=1}^{\infty} f(n).$$

This is easy for functions which only have finitely many non-zero terms: if $g$ is a function on $\mathbb{N}$ such that $g(n) = 0$ for $n > N$, then we can write

$$g = \sum_{n=1}^{N} g(n) \chi_{\{n\}}$$

so $g$ is simple and

$$\int_{\mathbb{N}} g \, \mathrm{d}\mu = \sum_{n=1}^{N} g(n).$$

Now, if $f$ is any non-negative function on $\mathbb{N}$, then for any $N \in \mathbb{N}$ we can define a function $g$ so $g(n) = f(n)$ if $n \leq N$ and $g(n) = 0$ for $n > N$. Certainly $0 \leq g \leq f$, so $\int_{\mathbb{N}} g \, \mathrm{d}\mu \leq \int_{\mathbb{N}} f \, \mathrm{d}\mu$ and hence

$$\sum_{n=1}^{N} f(n) \leq \int_{\mathbb{N}} f \, \mathrm{d}\mu$$

for all $N \in \mathbb{N}$. Taking the supremum over $N$ we see that

$$\sum_{n=1}^{\infty} f(n) \leq \int_{\mathbb{N}} f \, \mathrm{d}\mu \, .$$

If the LHS is $\infty$, then this implies that $\int_{\mathbb{N}} f \, \mathrm{d}\mu = \infty$, and we have nothing more to prove.

If the LHS is finite, then we still have to establish the reverse inequality. Suppose that $g$ is any non-negative simple function with $g \leq f$. Because $f$ is simple, it must be constant except for finitely many terms. But $0 \leq g(n) \leq f(n) \to 0$, so $g$ must be zero except for finitely many terms, say $g(n) = 0$ for $n > N$. Now we have

$$\int_{\mathbb{N}} g \, \mathrm{d}\mu = \sum_{n=1}^{N} g(n) \leq \sum_{n=1}^{N} f(n) \leq \sum_{n=1}^{\infty} f(n)$$

and taking the supremum over all $g$ gives the required inequality.

If we extend this example to a complex function $f$, it is not hard to check (**exercise**) that

$$\int_{\mathbb{N}} f \, \mathrm{d}\mu = \sum_{n=1}^{\infty} f(n)$$

provided that the series on the RHS is *absolutely* convergent.

3.5. **Example.** If $(S, \Sigma, \mu)$ is a measure space, we can define the space $L^p(S, \Sigma, \mu)$ for $1 \leq p < \infty$ in exactly the same way as we did with the Lebesgue integral: take the space of all $\mu$-measurable functions on $S$ such that $|f|^p$ is $\mu$-integrable and quotient by $\mu$-null functions. Similarly, $L^\infty(S, \Sigma, \mu)$ can be constructed in exactly the same way as for the Lebesgue integral. Everything that was previously stated about $L^p$ spaces carries over word for word to the new context.

This allows to bridge the gap between $\ell^p$ and $L^p$ spaces: we use the measure space $(\mathbb{N}, 2^{\mathbb{N}}, \mu)$ where $\mu$ is counting measure, and identify sequences $(a_n)$ with measurable functions on $\mathbb{N}$. The previous example shows that integration with respect to this measure is the same as absolute summation, so all the sums in the definition of $\ell^p$ turn into integrals such as

$$\sum_{n=1}^{\infty} |a_n|^p = \int_{\mathbb{N}} |a|^p \, \mathrm{d}\mu$$

In summary, we simply have $\ell^p = L^p(\mathbb{N}, 2^{\mathbb{N}}, \mu)$ so any results we prove about $L^p$ spaces over an abstract measure space apply equally to $L^p(S)$ for some $S \subseteq \mathbb{R}^N$ and $\ell^p$.

3.6**. Exercise.** Show that simple functions are dense in $L^p(S, \Sigma, \mu)$ for any measure space $(S, \Sigma, \mu)$ and any $p \in [1, \infty]$. *Suggestion: prove it for non-negative functions first, direct from the definition of the integral. The inequality $(x - y)^p \leq x^p - y^p$ if $0 \leq y \leq x$ and $p \in [1, \infty)$ is useful. The $p = \infty$ case is different but easier.*

The two important convergence theorems work in the context of general measure spaces exactly as they do in the Lesbesgue theory:

3.7**. Theorem** (Monotone Convergence Theorem (MCT))**.** *Let $(S, \Sigma, \mu)$ be a measure space and $(f_n)$ be a monotone sequence of $\mu$-integrable functions from $S$ to $\mathbb{K}$ such that $\int_S f_n \, \mathrm{d}\mu$ remains bounded as $n \to \infty$. Then there is a $\mu$-integrable function $f : S \to K$ such that $f_n \to f$ $\mu$-a.e. and $\int_S f_n \, \mathrm{d}\mu \to \int_S f \, \mathrm{d}\mu$.*

3.8**. Theorem** (Dominated Convergence Theorem (DCT))**.** *Let $(S, \Sigma, \mu)$ be a measure space and $(f_n)$ be a sequence of $\mu$-integrable functions from $S$ to $\mathbb{K}$. Suppose that there exists a $\mu$-integrable function $g : S \to \mathbb{R}$ such that $|f_n(t)| \leq g(t)$ for all $n \in \mathbb{N}$ and $\mu$-almost all $t \in S$ and that there exists a function $f : S \to \mathbb{K}$ such that $f_n(t) \to f(t)$ for $\mu$-almost all $t \in S$. Then $f$ is $\mu$-integrable and $\int_S f_n \, \mathrm{d}\mu \to \int_S f \, \mathrm{d}\mu$.*

We record for future use one application of the MCT:

3.9**. Lemma.** *Suppose $(S, \Sigma, \mu)$ is a measure space and $f$ is a non-negative $\mu$-integrable function. Then there is an increasing sequence $(f_n)$ of simple functions such that $f_n \to f$ a.e. and $\int_S f_n \, \mathrm{d}\mu \to \int_S f \, \mathrm{d}\mu$.*

*Proof.* By definition, there is a sequence of simple functions $(g_n)$ with $0 \leq g_n \leq f$ and $\int_S g_n \, \mathrm{d}\mu \to \int_S f \, \mathrm{d}\mu$. Let $f_n = \max(g_1, g_2, \ldots, g_n)$. This is an increasing sequence and $0 \leq g_n \leq f_n \leq f$. Integrating gives

$$0 \leq \int_S g_n \, \mathrm{d}\mu \leq \int_S f_n \, \mathrm{d}\mu \leq \int_S f \, \mathrm{d}\mu$$

so we also have $\int_S f_n \, \mathrm{d}\mu \to \int_S f \, \mathrm{d}\mu$. By the MCT, $f_n \to g$ a.e. where $g$ is some non-negative integrable function and $\int_S f_n \, \mathrm{d}\mu \to \int_S g \, \mathrm{d}\mu$. We now have $\int_S f \, \mathrm{d}\mu = \int_S g \, \mathrm{d}\mu$ and $g \leq f$ $\mu$-a.e., so $\int_S |f - g| \, \mathrm{d}\mu = 0$ and hence $g = f$ $\mu$-a.e. ☐

By the discussion in Example 3.5, the following result applies to $L^p(S)$ if $S \subseteq \mathbb{R}^N$ and to $\ell^p$.

3.10**. Theorem.** *If $(S, \Sigma, \mu)$ is a measure space, then $L^p(S, \Sigma, \mu)$ is a B-space.*

*Proof.* We abbreviate $L^p(S, \Sigma, \mu)$ to $L^p$ throughout, and largely suppress reference to $\mu$, saying 'a.e.' for '$\mu$-a.e.', etc.

By Theorem 1.4, it is sufficient to show that every absolutely convergent series in $L^p$ is convergent. Suppose then that

$$\sum_{k=1}^{\infty} \|f_k\|_p = C$$

and let

$$g_n = \sum_{k=1}^{n} |f_k|.$$

It follows from Minkowski's inequality that $\|g_n\|_p \le C$, or more explicitly

$$\int_S g_n^p \, \mathrm{d}\mu \le C^p.$$

We thus have an increasing sequence $(g_n^p)$ of integrable functions whose integrals are bounded above. By the monotone convergence theorem, there is a non-negative integrable function $h$ on $S$ such that $g_n^p \to h$ a.e. Let $g = h^{1/p}$, so $g_n \to g$ a.e. This tells us that for almost all $s \in S$, $\sum f_k(s)$ is absolutely convergent and hence convergent; thus, if

$$F_n = \sum_{k=1}^{n} f_k$$

then $F_n$ is almost everywhere convergent to some function $F$, say. Moreover, $|F_n| \le g_n \le g$ so $|F| \le g$ a.e. and hence $|F_n - F| \le 2g$. This gives us $|F_n - F|^p \le 2^p g^p$, the LHS converges a.e. to zero, and the RHS is integrable. By the dominated convergence theorem, $\int_S |F_n - F|^p \, \mathrm{d}\mu \to 0$ as $n \to \infty$, which is to say that $\|F_n - F\|_p \to 0$. This shows that $\sum f_k$ is convergent, so the space is complete. $\square$

3.11**. Exercise.** If $(S, \Sigma, \mu)$ is a measure space, show that $L^\infty(S, \Sigma, \mu)$ is a B-space.

We can also consider complex-valued measures, as long as we exclude the case of infinite measure.

3.12**. Definition.** Let $\Sigma$ be a $\sigma$-field on a set $S$. A function $\mu : \Sigma \to \mathbb{C}$ is called a *complex measure* if $\mu(\emptyset) = 0$ and whenever $(A_n)$ is a sequence of disjoint sets in $\Sigma$,

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

Note that no $\infty$ is allowed here, so the sequence on the RHS is always required to converge. In fact, since the sequence $(A_n)$ can be rearranged in any way without affecting the LHS, it follows from the Riemann rearrangement theorem that the sum on the RHS is absolutely convergent. If $\mu$ is actually real-valued, it is usually referred to as a *signed measure*.

We now give a brief description of one of the most important results in measure theory.

3.13**. Definition.** Suppose that $\Sigma$ is a $\sigma$-field on a set $S$ and that $\mu$ and $\nu$ are two (positive or complex) measures on $S$. We say that $\mu$ is *absolutely continuous (a.c.)* with respect to $\nu$, written $\mu \ll \nu$, if every $\nu$-null set is also $\mu$-null.

We say that $\mu$ is *concentrated* on a set $S' \in \Sigma$ if $\mu(A) = \mu(A \cap S')$ for all $A \in \Sigma$. The measures $\mu$ and $\nu$ are said to be *mutually singular*, written $\mu \perp \nu$, if there are disjoint sets $S', S'' \in \Sigma$ such that $\mu$ is concentrated on $S'$ and $\nu$ is concentrated on $S''$.

3.14**. Example.** Trivial example: the zero measure is absolutely continuous with respect to any measure, and mutually singular with any measure (because it is concentrated on $\emptyset$).

The simplest way to construct a non-trivial absolutely continuous measure is by integration. If $(S, \Sigma, \nu)$ is a measure space and $h$ a complex-valued $\nu$-integrable function on $S$, we can define

$$\mu(A) = \int_S h\chi_A \, \mathrm{d}\nu \, .$$

It is (**exercise**) an easy consequence of the monotone and dominated convergence theorem that this is indeed a complex measure, and it is immediate that it is a.c. with respect to $\nu$. In fact, in many examples every measure a.c. with respect to $\mu$ can be represented in this way: see Theorem 3.15 below.

The simplest non-trivial example of a pair of mutually singular measures is Lebesgue measure on $\mathbb{R}$ and a Dirac measure $\delta_x$ for some $x \in \mathbb{R}$. The Dirac measure is concentrated on $\{x\}$ and Lebesgue measure on $\mathbb{R} \setminus \{x\}$.

3.15**. Theorem** (Lebesgue Decomposition, Radon-Nikodym Theorem)**.** *Suppose that $\Sigma$ is a $\sigma$-field on a set $S$, that $\nu$ is a positive $\sigma$-finite measure on $\Sigma$ and that $\mu$ is a complex measure on $\Sigma$. Then there is a unique pair of complex measures $\mu_{ac}$ and $\mu_{sing}$ such that $\mu_{ac}$ is absolutely continous with respect to $\nu$, $\mu_{sing}$ and $\nu$ are mutually singular and $\mu = \mu_{ac} + \mu_{sing}$ (Lebesgue decomposition). Further (Radon-Nikodym Theorem), there exists a $\nu$-integrable function $h$ (the* Radon-Nikodym derivative *of $\mu_{ac}$ with respect to $\nu$), unique up to a $\nu$-null set, such that for all $A \in \Sigma$,*

$$\mu_{ac}(A) = \int_S h\chi_A \, \mathrm{d}\nu \, ,$$

*so we actually have*

$$\mu(A) = \int_S h\chi_A \, \mathrm{d}\nu \, + \mu_{sing}(A).$$

3.16**. Exercise.** The $\sigma$-finiteness hypothesis is important. Let $\mu$ be Lebesgue measure on $[0, 1]$ and $\nu$ be counting measure on the same $\sigma$-field. Show that $\nu$ has no Lebesgue decomposition relative to $\mu$, and that $\mu$ is a.c. with respect to $\nu$ but has no Radon-Nikodym derivative with respect to $\nu$.

## 4. Dual Spaces

4.1**. Definition.** If $X$ is a normed space over $\mathbb{K}$, a linear map $\phi : X \to \mathbb{K}$ is called a *(linear) functional*. We say that $\phi$ is *bounded* if there is a constant $C$ such that $|\phi(x)| \leq C\|x\|$ for all $x \in X$, in which case we let $\|\phi\|$ be the infimum of all such $C$. The *dual space* $X^*$ of $X$ is the space of all bounded linear functionals on $X$.

We often denote elements of the dual space by $x^*$, $y^*$ etc. This is purely a notational convention: if we write $x \in X$ and $x^* \in X^*$, this is not intended of itself to suggest any relationship between $x$ and $x^*$.

4.2**. Exercise.** If $X$ is a normed space, $X^*$ its dual, $x \in X$ and $x^* \in X^*$, show that:

(1) $|x^*(x)| \leq \|x^*\|\|x\|$;

(2) $\|x^*\| = \sup_{x\in X, \|x\|=1} |x^*(x)| = \sup_{x\in X, \|x\|\le 1} |x^*(x)| = \sup_{x\in X, x\neq 0} |x^*(x)|/\|x\|$.
(3) $\|\cdot\|$ is a norm on $X^*$;
(4) a linear function on $X$ is bounded if and only if it is continuous.

4.3**. Theorem.** *If $X$ is a normed space, then $X^*$ is a B-space.*

*Proof.* Suppose $(x_n^*)$ is a Cauchy sequence in $X^*$ and $x \in X$. Since $(x_n^*)$ is Cauchy, it is bounded; say $\|x_n^*\| < C$ for all $n$. For any $x \in X$ and any $m, n \in \mathbb{N}$ we have

$$|x_m^*(x) - x_n^*(x)| = |(x_m^* - x_n^*)x| \le \|x_m^* - x_n^*\|\|x\|.$$

It follows that for each $x$ the sequence $(x_n^*(x))$ is Cauchy in $\mathbb{K}$, hence convergent in $\mathbb{K}$. We can therefore define a functional $x^* : X \to \mathbb{K}$ by

$$x^*(x) = \lim_{n\to\infty} x_n^*(x).$$

It is easy to check (**exercise**) that $x^*$ is linear. To show that $x^*$ is bounded, we have

$$|x^*(x)| = \lim_{n\to\infty} |x_n^*(x)| \le C\|x\|$$

so $x^*$ is indeed bounded, with $\|x^*\| \le C$. □

In many of the B-spaces $X$ introduced above, we can find an explicit realisation of the dual, by which we mean a concretely described B-space isometrically isomorphic to $X^*$, with a formula for the duality pairing.

4.4**. Example.** The Hölder inequality gives us many examples of bounded linear functionals. Fix a measure space $(S, \Sigma, \mu)$ and $p \in (1, \infty)$ and let $q$ be conjugate to $p$. We abbreviate $L^p(S, \Sigma, \mu)$ and $L^q(S, \Sigma, \mu)$ to $L^p$ and $L^q$.

For any $y \in L^q$, define $\phi : L^p \to \mathbb{K}$ by

$$\phi(x) = \int_S xy \,\mathrm{d}\mu\,.$$

By Hölder's inequality, this integral exists and

$$|\phi(x)| \le \|x\|_p \|y\|_q.$$

Linearity is easy to check, so $\phi$ is a bounded linear functional on $\ell^p$ with $\|\phi\| \le \|y\|_q$. We can actually go further, and conclude that $\|\phi\| = \|y\|_q$. To see this, let $x$ be a function on $S$ such that for all $t \in S$, $|x(t)|^p = |y(t)|^q$ and $x(s)y(s) \ge 0$ (convince yourself that such an $x$ always exists, and is even unique). It is immediate that $x \in L^p$, and if you backtrack through the proof of Hölder's inequality you will see that all the inequalities in the proof are in fact equalities in this case (this is half of the solution to Exercise 2.7). For this particular $x$, then, we have $|\phi(x)| = \|x\|_p \|y\|_q$, so $\|\phi\| \ge \|y\|_q$ and hence $\|\phi\| = \|y\|_q$.

It turns out that we have not just constructed a class of examples here: in fact, we have a concrete 'formula' for the entire dual space.

Suppose initially that $\mu$ is a finite measure. We can define a function $\nu$ on $\Sigma$ by

$$\nu(E) = \phi(\chi_E).$$

It is straightforward to check (**exercise**) that this is a complex measure which is absolutely continuous with respect to $\mu$. By the Radon-Nikodym Theorem, there exists a $\mu$-integrable function $h$ such that
$$\nu(E) = \int_S h\chi_E \,\mathrm{d}\mu$$
so we have
$$\phi(\chi_E) = \int_S h\chi_E \,\mathrm{d}\mu\,.$$
It is now an immediate consequence of linearity that
$$\phi(x) = \int_S hx \,\mathrm{d}\mu \tag{*}$$
whenever $h$ is a simple function. We claim that this implies that $h \in L^q$. Assuming for the moment that this is true, it follows from the earlier part of the example that the RHS of (*) defines a bounded linear functional on the whole of $L^p$, so we have two two bounded linear functional which agree on the dense (exercise 3.6) subspace of simple functions. They must therefore agree everywhere, so (*) is true for all $x \in L^p$. Every functional in $(L^p)^*$ can thus be represented by integration against a function in $L^q$.

It is not hard to extend this to $\sigma$-finite measures. If $(S, \Sigma, \mu)$ is $\sigma$-finite, then we can write
$$S = \bigcup_{n\in\mathbb{N}} S_n$$
where the $S_n$ are pairwise disjoint and have finite measure. Each of the $S_n$ can be regarded as a measure space in its own right, with $\sigma$-field
$$\Sigma_n = \{A \cap S_n : A \in \Sigma\}$$
and measure the restriction of $\mu$. Any bounded linear functional on $L^p(S, \Sigma, \mu)$ gives rise to a a functional $\phi_n$ on each $L^p(S_n)$: if $x \in L^p(S_n)$, then extend $x$ to a function $x'$ defined on $S$ by making $x'(t) = x(t)$ if $t \in S_n$ and $x'(t) = 0$ if $t \in S \setminus S_n$, and let $\phi_n(x) = \phi(x')$. It is immediate that $\phi_n$ is bounded with $\|\phi_n\| \le \|\phi\|$, so we have $h_n \in L^q(S_n)$ such that $\phi_n(x) = \int_{S_n} xh_n \,\mathrm{d}\mu$. Now define $h$ by $h(t) = h_n(t)$ if $t \in S_n$. It is easy to check that $h \in L^q(S)$ and that $\phi(x) = \int_S hx \,\mathrm{d}\mu$ for all $x \in L^p(S)$. Details left as an **exercise**.

To summarise: for any $\sigma$-finite measure space $(S, \Sigma, \mu)$, we have constructed an isometric isomorphism between $(L^p(S, \Sigma, \mu))^*$ and $L^q(S, \Sigma, \mu)$, with an explicit formula describing the duality action.

In fact, the hypothesis that $\mu$ is $\sigma$-finite is unnecessary: see, for example, §15 of Hewitt and Stromberg [3].

It remains to show:

4.5. **Lemma.** *Let $(S, \Sigma, \mu)$ be a finite measure space and suppose that $h$ is an integrable function from $S$ to $\mathbb{K}$ and there exists a constant $C$ such that*
$$\left| \int_S xh \,\mathrm{d}\mu \right| \le C\|x\|_p$$

*for all simple functions $x$. Then $h \in L^q$ and $\|h\| \leq C$.*

*Proof.* Assume WLOG that $h$ is non-negative. Since $h$ is integrable, there exists (Lemma 3.9) an increasing sequence $(h_n)$ of simple functions such that $h_n \to h$ a.e. Let $x_n = h_n^{q/p}$. By hypothesis,

$$\left| \int_S x_n h \,\mathrm{d}\mu \right| \leq C \|x_n\|_p$$

which expands to

$$\int_S h_n^{q/p} h \,\mathrm{d}\mu \leq C \left( \int_S h_n^q \,\mathrm{d}\mu \right)^{1/p}.$$

Now, $h_n \leq h$, so we have

$$\int_S h_n^{1+q/p} \,\mathrm{d}\mu \leq C \left( \int_S h_n^q \,\mathrm{d}\mu \right)^{1/p}.$$

But $1 + q/p = q$ and $1 - 1/p = 1/q$, so

$$\left( \int_S h_n^q \,\mathrm{d}\mu \right)^{1/q} \leq C.$$

Finally, $h_n^q$ is an increasing sequence of functions converging to $h^q$ and $\int_S h_n^q \,\mathrm{d}\mu \leq C^q$. By the MCT, $h^q$ is integrable and $\int_S h^q \,\mathrm{d}\mu \leq C^q$, so $h \in L^q$ and $\|h\|_q \leq C$. □

4.6. **Exercise.** Show that if $(S, \Sigma, \mu)$ is $\sigma$-finite then $(L^1(S, \Sigma, \mu))^*$ can be identified with $L^\infty(S, \Sigma, \mu)$ in a similar way. Give an example to show that the $\sigma$-finiteness hypothesis is not redundant in this case (although it can be weakened a little by slightly changing the definition of $L^\infty$ for non-$\sigma$-finite spaces: see Theorem 20.20 in Hewitt and Stromberg [3, Theorem 20.20]).

4.7. **Exercise.** In the special case $\ell^p$ we can prove this without using any measure theory at all. Instead of using the Radon-Nikodym Theorem, we proceed as follows: if we have a functional $\phi$ on $\ell^p$ then for any $x \in \ell^p$ we have

$$\phi(x) = \phi\left( \sum_{n=1}^\infty x_n e^{(n)} \right) = \sum_{n=1}^\infty x_n \phi(e^{(n)}) = \sum_{n=1}^\infty x_n h_n.$$

Write out a complete proof of $\ell^p$-$\ell^q$ duality using this approach.

4.8. **Exercise.** Use the same approach to show that $c_0^*$ can be identified with $\ell^1$ and $(\ell^1)^*$ with $\ell^\infty$. (The space $c_0$ was defined in Exercise 2.20).

4.9. **Example.** If $q$ is the conjugate of $p \in (1, \infty)$, then $p$ is the conjugate of $q$. It follows that $L^q$ is (isomorphically isomorphic to) the dual of $L^p$ as well as vice-versa. This is an important property which we shall come back to later when we talk about reflexive spaces in Section 6.

However, not all B-spaces behave this way. For example, you saw in Exercise 4.8 that the dual of the space $c_0$ can be identified with $\ell^1$ and the dual of $\ell^1$ with $\ell^\infty$. If these spaces behaved like the other $\ell^p$ spaces then $c_0$ would be isometrically isomorphic to $\ell^\infty$, which it is not.

Probably the easiest way to see this is to note that $c_0$ is separable, but $\ell^\infty$ is not. We can exhibit a countable dense subset of $c_0$ easily enough: the set of all rational sequences with only finitely many non-zero terms is one example. On the other hand, we can associate with each subset $A$ of $\mathbb{N}$ the sequence whose $n$th term is $1$ if $n \in A$ and $0$ if not. This gives an uncountable family of sequences in $\ell^\infty$, each distant $1$ from all the others, which shows that $\ell^\infty$ is inseparable.

## 5. Extensions of Linear Functionals

In order to uniquely define a bounded linear functional, we do not need to specify its values everywhere: a dense subspace is enough (in fact, by linearity, a set with dense span is enough).

5.1. **Lemma.** *If $Y$ is a dense subspace of a normed space $X$ and $\phi$ is a bounded linear functional on $Y$, then there is a unique bounded linear functional $\tilde{\phi}$ on $X$ which agrees with $\phi$ on $Y$. This extension is such that $\|\tilde{\phi}\| = \|\phi\|$.*

*Proof.* Note first that if such an extension $\tilde{\phi}$ exists, then for any $x \in X$ we can choose a sequence $(y_n)$ in $Y$ converging to $X$ and by continuity of $\tilde{\phi}$,

$$\tilde{\phi}(x) = \lim_{n\to\infty} \tilde{\phi}(y_n) = \lim_{n\to\infty} \phi(y_n). \tag{*}$$

This at once settles uniqueness, and shows us how to construct the extension. To prove the required result, we have to show:

(1) that if $(y_n)$ is convergent then $(\phi(y_n))$ is convergent;
(2) that if $y_n \to x$ and and $z_n \to x$ then $\phi(y_n)$ and $\phi(z_n)$ converge to the same limit, so (*) can be used to define a functional $\tilde{\phi}$;
(3) that $\tilde{\phi}$ so defined is linear and bounded with $\|\tilde{\phi}\| = \|\phi\|$.

This is all routine and is left as an **exercise**. □

We now approach the Hahn-Banach Theorem, which is the basic result which makes dual spaces useful. It states that if we have a bounded functional $\phi$ defined on a subspace of a normed space $X$, then $\phi$ can be extended to the whole of $X$ without increasing its norm. This turns out to be a very powerful tool, allowing us to construct functionals meeting various constraints. The argument is inductive in nature: we extend the functional one dimension at a time. Unusually, we need to treat the real and complex cases separately.

5.2. **Lemma.** *Suppose that $X$ is a real normed space, $Y$ is a proper subspace of $X$, $\phi$ is a bounded linear functional on $Y$ and $x_0 \in X$. Let $Z$ be the subspace spanned by $Y$ and $x_0$. Then there is a bounded linear functional $\tilde{\phi}$ on $Z$ which agrees with $\phi$ on $Y$ and is such that $\|\tilde{\phi}\| = \|\phi\|$.*

*Note that the norms are defined with respect to the stated domains of the functionals, so $\|\phi\| = \sup_{y\in Y, \|y\|=1} |\phi(y)|$ and $\|\tilde{\phi}\| = \sup_{z\in Z, \|z\|=1} |\phi(z)|$.*

*Proof.* Trivial case: if $x_0 \in Y$, then $Z = Y$ and we can just let $\tilde{\phi} = \phi$. Assume therefore that $x_0 \notin Y$, so any $z \in Z$ can be uniquely decomposed as $z = y + \lambda x_0$ where $y \in Y$ and $\lambda \in \mathbb{R}$, giving $\tilde{\phi}(z) = \phi(y) + \lambda\tilde{\phi}(x_0)$. The construction of $\tilde{\phi}$ is thus equivalent to the choice of one scalar $\alpha = \tilde{\phi}(x_0)$. The norm is given by

$$\|\tilde{\phi}\| = \sup_{y\in Y, \lambda\in\mathbb{R}} \frac{|\phi(y) + \lambda\alpha|}{\|y + \lambda x_0\|}.$$

If we divide top and bottom by $-\lambda$ and use the linearity of $\phi$, this gives

$$||\tilde{\phi}|| = \sup_{y \in Y, \lambda \in \mathbb{R}} \frac{|\alpha - \phi(y/\lambda)|}{||x_0 - y/\lambda||}$$

and we can now see that $\lambda$ is irrelevant: $y/\lambda$ is just an arbitrary element of $y$, so in fact

$$||\tilde{\phi}|| = \sup_{y \in Y} \frac{|\alpha - \phi(y)|}{||x_0 - y||}.$$

It is apparent that $||\tilde{\phi}|| \geq ||\phi||$, so we seek an $\alpha$ such that $||\tilde{\phi}|| \leq ||\phi||$, or equivalently such that

$$|\alpha - \phi(y)| \leq ||\phi|| ||x_0 - y||$$

for all $y \in Y$. This is equivalent to

$$\phi(y) - ||\phi|| ||x_0 - y|| \leq \alpha \leq \phi(y) + ||\phi|| ||x_0 - y||$$

for all $y \in Y$ which is in turn equivalent to

$$\phi(y_1) - ||\phi|| ||x_0 - y_1|| \leq \phi(y_2) + ||\phi|| ||x_0 - y_2||$$

for all $y_1, y_2 \in Y$. But this inequality is easy to prove:

$$\phi(y_1) - \phi(y_2) = \phi(y_1 - y_2) \leq ||\phi|| ||y_1 - y_2|| \leq ||\phi|| ||y_1 - x_0|| + ||\phi|| ||x_0 - y_2||$$

and rearrange. This proves that a suitable $\alpha$ exists, so $||\tilde{\phi}|| = ||\phi||$. □

5.3. **Exercise.** Worry about various potential divisions by zero in the above argument. At what point is the completeness of $\mathbb{R}$ used?

5.4. **Corollary** (Hahn-Banach Theorem for real, separable spaces). *Suppose that $X$ is a separable real normed space, $Y$ is a subspace of $X$ and $\phi$ is a bounded linear functional on $Y$. Then there is a bounded linear functional $\tilde{\phi}$ on $X$ which agrees with $\phi$ on $Y$ and is such that $||\tilde{\phi}|| = ||\phi||$.*

*Proof.* Since $X$ is separable, we can find a sequence $(x_n)$ which is dense in $X$. Let $Y_0 = Y$ and for $n \in \mathbb{N}$ let $Y_n$ be the subspace spanned by $Y$ and the vectors $x_1, \ldots, x_n$. For each $n \in \mathbb{N}$, $Y_n$ is spanned by $Y_{n-1}$ and $x_n$, so we can extend $\phi$ from $Y$ to $Y_1$ to $Y_2$ and ultimately to the subspace spanned by $Y$ and all the $x_n$. The norms of all these extensions are $||\phi||$. This subspace is dense in $X$ by the choice of the sequence $(x_n)$, so by Lemma 5.1 we can finally extend $\phi$ to $\tilde{\phi}$ on the whole of $X$, with $||\tilde{\phi}|| = ||\phi||$. □

If $X$ is not separable, then this procedure fails and we have to use some slightly heavier technology: Zorn's Lemma, a set-theoretic result equivalent to the Axiom of Choice (there is an accessible discussion of this and related results in Halmos [2] §§15–17).

5.5. **Definition.** A *partially ordered set* is a set $A$ with a binary operation $\prec$ satisfying

(1) $a \prec a$ (reflexive);
(2) $a \prec b$ and $b \prec c$ implies $a \prec c$ (transitive);
(3) $a \prec b$ and $b \prec a$ implies $a = b$ (anti-symmetric).

A subset $B$ of $A$ is called *linearly ordered* if for any $a, b \in B$ we have either $a \prec b$ or $b \prec a$.

A point $a \in A$ is called an *upper bound* for $B \subseteq A$ if $b \prec a$ for all $b \in B$.

A *maximal element* of $A$ is an element $a$ with the property that if $a \prec b$ then $a = b$. Note that this is *not* the same as saying that $a$ is an upper bound for $A$! Any upper bound for $A$ is maximal, but not vice-versa.

Obvious examples are $\mathbb{R}$ or $\mathbb{Z}$ or $\mathbb{N}$ with $\leq$ or $\geq$, or the power set of some set, with $\subseteq$ or $\supseteq$.

5.6. **Theorem** (Zorn's Lemma). *If $A$ is a partially ordered set with the property that every linearly ordered subset of $A$ has an upper bound, then every linearly ordered subset of $A$ has an upper bound which is a maximal element of $A$.*

We can now remove the word 'separable' from the hypotheses of the Hahn-Banach Theorem.

5.7. **Corollary** (Hahn-Banach Theorem for arbitrary real spaces). *Suppose that $X$ is a real normed space, $Y$ is a subspace of $X$ and $\phi$ is a bounded linear functional on $Y$. Then there is a bounded linear functional $\tilde{\phi}$ on $X$ which agrees with $\phi$ on $Y$ and is such that $\|\tilde{\phi}\| = \|\phi\|$.*

*Proof.* Let $\mathscr{Z}$ be the set of all subspaces $Z$ of $X$ such that $Z$ contains $Y$ and there exists a suitable (i.e. linear with norm $\|\phi\|$) extension of $\phi$ to $Z$, and consider the set $\mathscr{S}$ of all pairs $(Z, \tilde{\phi})$ where $Z \in \mathscr{Z}$ and $\tilde{\phi}$ is a suitable extension of $\phi$ to $Z$. Order this set of pairs by the relationship $(Z_1, \tilde{\phi}_1) \prec (Z_2, \tilde{\phi}_2)$ if $Z_1 \subseteq Z_2$ and $\tilde{\phi}_2$ agrees with $\tilde{\phi}_1$ on $Z_1$. Now, if we take a linearly ordered subset of $\mathscr{S}$, say

$$\{(Z_\alpha, \tilde{\phi}_\alpha) : \alpha \in A\}$$

for some index set $A$, then we can define $Z = \bigcup_{\alpha \in A} Z_\alpha$ and the linear ordering guarantees that this is a subspace. Moreover, if $z \in Z$ then $z \in Z_\alpha$ for some $\alpha \in A$ and we can define $\tilde{\phi}(z) = \tilde{\phi}_\alpha(z)$; the linear ordering guarantees that this value does not depend on $\alpha$. Moreover,

$$|\tilde{\phi}(z)| = |\tilde{\phi}_\alpha(z)| \leq \|\tilde{\phi}_\alpha\| \|z\| = \|\phi\| \|z\|$$

so $\tilde{\phi}$ is a suitable extension of $\phi$ to $Z$ and $(Z, \tilde{\phi})$ is an upper bound for the linearly ordered set we started with.

Now, Zorn's Lemma tells us that every linearly ordered subset of $\mathscr{S}$ has an upper bound which is maximal. All we actually need from this is that there is a maximal element of $\mathscr{S}$, say $(Z, \tilde{\phi})$. If $Z$ were a proper subspace of $X$, then Lemma 5.2 would give us another extension of $\tilde{\phi}$, contradicting maximality. If follows that $Z = X$ and that $\tilde{\phi}$ is a suitable extension. $\square$

5.8. **Exercise.** Check the details in the construction of the upper bound for $\{(Z_\alpha, \tilde{\phi}_\alpha) : \alpha \in A\}$.

With a little algebraic juggling, we can do the same for complex space.

5.9. **Corollary** (Hahn-Banach Theorem for arbitrary complex spaces). *Suppose that $X$ is a complex normed space, $Y$ is a subspace of $X$ and $\phi$ is a bounded linear functional on $Y$. Then there is a bounded linear functional $\tilde{\phi}$ on $X$ which agrees with $\phi$ on $Y$ and is such that $\|\tilde{\phi}\| = \|\phi\|$.*

*Proof.* We can regard $X$ as a real normed space and $\phi$ as a real linear functional on $Y$. The real Hahn-Banach Theorem tells us that there exists an extension $\psi$ of $\phi$ from $Y$ to $X$ which is real linear; that is, it satisfies $\psi(x_1 + x_2) = \psi(x_1) + \psi(x_2)$ for all $x_1, x_2 \in X$ and $\psi(\lambda x) = \lambda\psi(x)$ for $x \in X$ and $\lambda$ real. Now define

$$\tilde{\phi}(x) = \psi(x) - \mathrm{i}\psi(\mathrm{i}x).$$

A little algebra shows that $\tilde{\phi}$ is complex linear. Moreover, $\mathrm{Re}\,\tilde{\phi} = \psi$, and it is not hard to deduce from this (**exercise**) that $\|\tilde{\phi}\| = \|\psi\| = \|\phi\|$. □

Here is a typical application of the Hahn-Banach Theorem:

5.10**. Corollary.** *If $X$ is a normed space, $Y$ is a subspace of $X$ which is not dense in $X$ and $x \in X$ with $\mathrm{dist}(x, Y) = d$ then there is a functional $x^* \in X^*$ such that $\|x^*\| = 1$, $x^*(x) = d$ and $x^*(y) = 0$ for $y \in Y$, and indeed*

$$\mathrm{dist}(x, Y) = \max_{\substack{x^* \in X^* \\ x^* \equiv 0 \text{ on } Y}} |x^*(x)|.$$

*Proof.* Assume initially that $d > 0$ and let $Z$ be the subspace spanned by $x$ and $Y$. Any vector $z \in Z$ can be uniquely written as $z = y + \lambda x$ where $y \in Y$ and $\lambda \in \mathbb{K}$. Define a functional on $Z$ by $\phi(y + \lambda x) = \lambda d$, so $\phi(x) = d$ and $\phi$ is zero on $Y$. It is easy to check that $\phi$ is linear, and

$$\|\phi\| = \sup_{y \in Y, \lambda \in \mathbb{K}} \frac{|\lambda d|}{\|y + \lambda x\|}$$

Dividing top and bottom by $-\lambda$ and noting that $y/\lambda$ is just an arbitrary member of $Y$ (the same trick as we used in Lemma 5.2), we see that

$$\|\phi\| = \sup_{y \in Y} \frac{d}{\|x - y\|} = 1$$

by definition of $d$. We can now extend $\phi$ to $x^*$ with $\|x^*\| = 1$ using the Hahn-Banach Theorem. To deduce the identity for $\mathrm{dist}(x, Y)$ notice that if $x^* \in X^*$, $x^*(y) = 0$ for $y \in Y$ and $\|x^*\| = 1$ then for any $y \in Y$,

$$|x^*(x)| = |x^*(x - y)| \leq \|x^*\|\|x - y\|$$

and take the inf over $y \in Y$ to give $|x^*(x)| \leq \mathrm{dist}(x, Y)$.

The case $d = 0$ is now trivial, for we then have $x \in \overline{Y}$ so if $x^*$ is zero on $Y$, then $x^*(x) = 0$. Moreover, functionals of norm 1 which are zero on $Y$ exist, because we can apply the first part to $Y$ and any point of $X \setminus \overline{Y}$, which is non-empty since $Y$ is not dense in $X$ by hypothesis. □

In the special case $Y = \{0\}$ we have $d = \|x\|$ and so:

5.11**. Corollary.** *Let $X$ be a normed space and $x \in X$. Then there exists $x^* \in X^*$ such that $\|x^*\| = 1$ and $x^*(x) = \|x\|$, and*

$$\|x\| = \max_{\substack{x^* \in X^* \\ \|x^*\| = 1}} |x^*(x)|.$$

This formula is a companion to

$$\|x^*\| = \sup_{\|x\| = 1} |x^*(x)|.$$

This kind of result is why we speak of 'duality'. One immediate consequence is that $X^*$ *separates points*: if $x^*(x) = x^*(y)$ for all $x^* \in X^*$, then $x = y$. Note that the formula for $\|x^*\|$ has a $\sup$, and that for $\|x\|$ has a $\max$. The $\sup$ cannot in general be replaced by a $\max$, as you will see in the following exercise (although in some spaces it can: see Section 6). Also, the $x^*$ constructed in the theorem may or may not be unique.

5.12**. Exercise.** You have seen (Exercise 4.8) that $c_0^*$ can be identified with $\ell^1$. Use this identification to give an example of a functional $x^* \in c_0^*$ with $\|x^*\| = 1$ such that $|x^*(x)| < \|x^*\|$ for all $x \in c_0$ with $\|x\| = 1$. Deduce that $c_0$ is not isometrically isomorphic to the dual of any normed space; in particular, $c_0$ is not isomorphic to $\ell^\infty$. (cf. Example 4.9).

Show that if $x \in X = L^p(S, \Sigma, \mu)$ with $p \in (1, \infty)$ and $(S, \Sigma, \mu)$ some $\sigma$-finite measure space, then there exists exactly one $x^* \in X^*$ with $\|x^*\| = 1$ and $x^*(x) = \|x\|$, but that this is not true if $p = 1$ or $p = \infty$ (except in the degenerate case where $\Sigma$ does not contain two disjoint sets of non-zero measure!).

5.13**. Corollary.** *If $X$ is a B-space and $X^*$ is separable, then $X$ is separable.*

*Proof.* Choose a countable family $\{x_n^* : n \in \mathbb{N}\}$ which is dense in the set $\{x^* \in X^* : \|x^*\| = 1\}$. For each $n$ there exists $x_n \in X$ with $\|x_n\| = 1$ and $|x_n^*(x_n)| \geq 1/2$. Let $Y$ be the subspace spanned by the $x_n$. We claim that $\overline{Y} = X$. If not, then there would be a point in $X$ at a positive distance from $Y$ and hence by Corollary 5.10 a functional $x^* \in X^*$ with $\|x^*\| = 1$ which is identically zero on $Y$. But

$$\frac{1}{2} \leq |x_n^*(x_n)| = \|(x_n^* - x^*)x_n\| \leq \|x_n^* - x^*\|\|x_n\| = \|x_n^* - x^*\|.$$

This contradicts the density of the $(x_n^*)$ in the unit sphere in $X^*$. We can therefore conclude that $Y$ is dense in $X$. But the set of (finite) rational linear combinations of the $(x_n)$ is countable and dense in $Y$, hence dense in $X$. This shows that $X$ is separable. □

## 6. Reflexivity

There is a general mechanism by which elements of $X$ can act as functionals on $X^*$.

6.1**. Definition.** Let $X$ be a normed space and $X^*$ its dual. Denote by $X^{**}$ the dual of $X^*$, also known as the *bidual* of $X$. Define a map $J : X \to X^{**}$ by

$$(Jx)(x^*) = x^*(x).$$

Corollary 5.11 tells us that this map is an isometry, so in particular it is injective. $X$ is said to be *reflexive* if this map is surjective (which implies that $X$ is isometrically isomorphic to its bidual).

6.2**. Exercise.** Show that if $X$ is a B-space then $J(X)$ is closed in $X^{**}$, and that if $X$ is reflexive then $X$ is a B-space. Show also that if $X$ is not complete then the closure of $J(X)$ in $X^{**}$ is a B-space in which $X$ is densely isometrically embedded (this is another way to construct the completion of $X$, instead of using equivalence classes of Cauchy sequences).

Note that although every reflexive space is isometrically isomorphic to its bidual, the converse is *not* true (although counterexamples are not easy to come by: see James [4]). The definition of reflexivity is that the particular map $J$ is an isomorphism, not just that an isomorphism exists.

6.3. **Example.** If $p \in (1, \infty)$ and $(S, \Sigma, \mu)$ is a $\sigma$-finite measure space then $L^p(S, \Sigma, \mu)$ is reflexive. We have already seen that this space is isometric to its bidual, but this is not enough: we need to check that the map $J$ from Definition 6.1 is an isomorphism, i.e. is surjective. This is a little fiddly, but not difficult.

Abbreviate $L^p(S, \Sigma, \mu)$ to $L^p$ and $L^q(S, \Sigma, \mu)$ to $L^q$ as usual, and let $q$ be the conjugate index to $p$. Let $T$ be the isometric isomorphism from $(L^p)^*$ to $L^q$ constructed in Example 4.4, which satisfies

$$x^*(y) = \int_S yTx^* \,\mathrm{d}\mu \tag{1}$$

and choose $x^{**} \in (L^p)^{**}$. We have to construct $x \in L^p$ such that $x^{**} = Jx$, or equivalently $x^{**}(x^*) = x^*(x)$ for all $x^* \in X^*$. We begin by shifting focus from $(L^p)^*$ to its concrete representation $L^q$: define a functional on $\ell^q$ by

$$\phi(z) = x^{**}(T^{-1}z). \tag{2}$$

By Example 4.4 again, but this time with $p$ and $q$ interchanged, there exists $x \in L^p$ such that

$$\phi(z) = \int_S zx \,\mathrm{d}\mu \tag{3}$$

for all $z \in L^q$. Now, for any $x^* \in (L^p)^*$,

$$x^{**}(x^*) = \phi(Tx^*) = \int_S xTx^* \,\mathrm{d}\mu = x^*(x) = (Jx)(x^*)$$

using (2), (3) and (1) in that order. This shows that $Jx = x^{**}$, so $J$ is indeed surjective.

Roughly speaking, the extra ingredient that makes this work is that the formula representing the duality action, $\int_S xy \,\mathrm{d}\mu$, is formally symmetric in $x$ and $y$. If you follow $x$, you will see that it starts off representing a vector in a dual space acting on something else, but ends up representing a vector being acted on by something in a dual space.

6.4. **Exercise.** The Riesz-Fréchet Theorem states that if $H$ is a Hilbert space and $\phi \in H^*$ then there exists $y \in H$ such that $\phi(x) = \langle x, y \rangle$ for all $x \in H$. Use this to show that any Hilbert space is reflexive.

6.5. **Theorem.** *A B-space $X$ is reflexive if and only if $X^*$ is reflexive.*

*Proof.* Throughout the proof, let $J_0$ be the canonical mapping from $X$ to $X^{**}$ and $J_1$ the canonical mapping from $X^*$ to $X^{***}$.

Suppose first that $X$ is reflexive, so $J_0$ is surjective. We have to show that $J_1$ is surjective. Given $x^{***} \in X^{***}$, define $x^* \in X^*$ by $x^*(x) = x^{***}(J_0x)$. We claim that $x^{***} = J_1x^*$. To see this, note that

$$(J_1x^*)(J_0x) = (J_0x)(x^*) = x^*(x) = x^{***}(J_0x)$$

using the definitions of $J_1$, $J_0$ and $x^*$. This shows that $J_1x^*$ and $x^{***}$ agree on the range of $J_0$. But $J_0$ is surjective, so they agree everywhere, which is to say that $x^{***} = J_1x^*$. This shows that $J_1$ is surjective, so $X^*$ is reflexive.

Now suppose that $X^*$ is reflexive, and suppose for a contradiction that $X$ is not reflexive. Then $J_0(X)$ is a closed, proper subspace of $X^{**}$, so by Corollary 5.10 there exists $x^{***} \in X^{***}$ such that $x^{***}(J_0 x) = 0$ for all $x \in X$, and $\|x^{***}\| = 1$. Since $X^*$ is reflexive, $x^{***} = J_1 x^*$ for some $x^* \in X^*$. We now have $(J_1 x^*)(J_0 x) = 0$ for all $x \in X$, and $\|x^*\| = 1$. But the first of these translates to $x^*(x) = 0$ for all $x \in X$, so $x^* = 0$, contradicting $\|x^*\| = 1$. This contradiction shows that $X$ is reflexive. $\square$

6.6. **Example.** We have seen (Exercise 4.8) that the dual of $c_0$ is (can be identified with) $\ell^1$ and the dual of $\ell^1$ is $\ell^\infty$, but (Exercise 5.12) that $c_0$ is not the dual of anything. This certainly shows that $c_0$ is not reflexive, from which it now follows that neither $\ell^1$ nor $\ell^\infty$ is reflexive.

This means that there must be a functional $\phi$ on $\ell^\infty$ which cannot be represented in the form

$$\phi(x) = \sum_{j=1}^{\infty} x_j y_j \tag{*}$$

for any $y \in \ell^1$ (in fact, any $y$ at all, for if this formula were to define a bounded functional on $\ell^\infty$, then $y$ would have to be in $\ell^1$; the proof of this is similar to the proof of Lemma 4.5).

We can (sort of) construct such a functional using the Hahn-Banach Theorem. Consider the subspace $c$ of $\ell^\infty$ consisting of all convergent sequences, and define $\phi$ on $c$ by

$$\phi(x) = \lim_{j\to\infty} x_j.$$

It is easy to check (**exercise**) that this is a bounded linear functional with $\|\phi\| = 1$, so by the Hahn-Banach Theorem it can be extended to a functional $\tilde{\phi}$ on the whole of $\ell^\infty$ with $\|\tilde{\phi}\| = 1$. But $\phi(e^{(n)}) = 0$ so $\tilde{\phi}(e^{(n)}) = 0$ as well, and if $\tilde{\phi}$ could be represented by a formula like (*), then we would have $y_n = \tilde{\phi}(e^{(n)})$, so $\tilde{\phi}$ would be identically zero, contradicting $\|\tilde{\phi}\| = 1$. (The existence of such a functional is another proof that $\ell^1$ is not reflexive.)

6.7. **Exercise.** Let $S$ be a compact subset of $\mathbb{R}^N$. Then the space of continuous functions $C(S)$ is a subspace of $L^\infty(S)$. Choose $s \in S$ and define a functional on $C(S)$ by $\phi(x) = x(s)$. Show that this is a bounded linear functional on $C(S)$ with $\|\phi\| = 1$, but that no extension $\tilde{\phi}$ of $\phi$ to $L^\infty(S)$ can be represented in the form

$$\tilde{\phi}(x) = \int_S xy$$

for any $y \in L^1(S)$. Deduce that neither $L^1(S)$ nor $L^\infty(S)$ is reflexive.

6.8. **Example.** Another way of obtaining these results: both $\ell^1$ and $L^1$ are separable, but their duals $\ell^\infty$ nor $L^\infty$ are inseparable. It follows from Corollary 5.13 that none of these spaces are reflexive.

## 7. Some Applications

We conclude with a couple of examples showing how we can obtain information about a B-space by using its dual.

7.1. **Theorem** (Riesz's Lemma). *Let $Y$ be a proper closed subspace of a normed B-space $X$. Then*

$$\sup_{\substack{x\in X\\ \|x\|=1}} \operatorname{dist}(x,Y)=1.$$

*In particular, for any $\varepsilon>0$ there exists $x\in X$ with $\|x\|=1$ such that $\|x-y\|>1-\varepsilon$ for all $y\in Y$.*

*Proof.* We know (Corollary 5.10) that for any $x\in X$,

$$\operatorname{dist}(x,Y)=\sup_{\substack{x^*\in X^*\\ \|x^*\|=1\\ x^*\equiv 0 \text{ on } Y}} |x^*(x)|$$

We thus have

$$\sup_{\substack{x\in X\\ \|x\|=1}} \operatorname{dist}(x,Y)=\sup_{\substack{x\in X\\ \|x\|=1}}\ \sup_{\substack{x^*\in X^*\\ \|x^*\|=1\\ x^*\equiv 0 \text{ on } Y}} |x^*(x)|.$$

Exchanging the order of the suprema gives us

$$\sup_{\substack{x^*\in X^*\\ \|x^*\|=1\\ x^*\equiv 0 \text{ on } Y}}\ \sup_{\substack{x\in X\\ \|x\|=1}} |x^*(x)|.$$

But the inner $\sup$ is now just $\|x^*\|$ which is $1$. □

Note that there does not always exist a point $x\in X$ with $\|x\|=1$ and $\operatorname{dist}(x,Y)=1$.

7.2. **Example.** Let $X=\{x\in C([0,1],\mathbb{R}) : x(0)=0\}$ with the supremum norm and

$$Y=\left\{y\in X : \int_0^1 y=0\right\}.$$

Then there does not exist $x_0\in X$ with $\|x_0\|=1$ and $\operatorname{dist}(x_0,Y)=1$. If there did, then for any $x\in X\setminus Y$ we could let $c=\left(\int_0^1 x_0\right)/\left(\int_0^1 x\right)$, so $y=x_0-cx$ would be an element of $Y$. But then we would have $\|y-x_0\|\geq 1$, so $|c|\|x\|\geq 1$, or explicitly

$$\left|\int_0^1 x\right|\leq\left|\int_0^1 x_0\right|\|x\|$$

for all $x\in X$. Now, if we let $x_n(t)=t^{1/n}$ then $\|x_n\|=1$ and $\int_0^1 x_n\to 1$ as $n\to\infty$. We must therefore have

$$\left|\int_0^1 x_0\right|\geq 1.$$

But $x_0$ is continuous, $x_0(0)=0$ and $|x_0(t)|\leq 1$ for all $t$, from which is follows that

$$\left|\int_0^1 x_0\right|<1.$$

This contradiction shows that no element $x_0$ with the stated property exists.

7.3. **Corollary.** *If $X$ is a normed space, then the unit ball $B = \{x \in X : \|x\| \le 1\}$ is compact if and only if $X$ is finite-dimensional.*

*Proof.* If $X$ is finite-dimensional, then $B$ is compact by the Heine-Borel Theorem and Theorem 1.7. If $X$ is infinite-dimensional, then we can use Riesz's Lemma to construct a sequence as follows: first choose any $x_1$ with $\|x_1\| = 1$ and let $X_1$ be the 1-dimensional subspace spanned by $x_1$. Then choose $x_2$ with $\|x_2\| = 1$ and $\operatorname{dist}(x_2, X_1) > 1/2$ and let $X_2$ be the subspace spanned by $x_1$ and $x_2$. Continuing in this way gives a sequence $(x_n)$ such that $\|x_n\| = 1$ for all $n$ and $\|x_n - x_m\| > 1/2$ whenever $m \ne n$. This sequence cannot have a convergent subsequence, so $B$ is not compact (the infinite-dimensionality is used here to guarantee that $X_n$ is always a proper subspace of $X$). □

In this construction, we arbitrarily used $\varepsilon = 1/2$ in Riesz's Lemma. We could have used other values of $\varepsilon$ to get $\|x_m - x_n\| > 1 - \varepsilon$, but we could never have got to $\|x_m - x_n\| \ge 1$. In some cases much larger distances are possible; for example, an orthonormal sequence $(e_n)$ in an inner product space has $\|e_m - e_n\| = \sqrt{2}$ whenever $m \ne n$. The following nice argument appears in Diestel [1] (An Afterthought to Riesz's Theorem, end of Chapter I, including credits for the theorem and for this proof).

7.4. **Theorem.** *If $X$ is an infinite-dimensional normed space then there is a sequence $(x_n)$ in $X$ with $\|x_n\| = 1$ for all $n$ such that $\|x_m - x_n\| > 1$ whenever $m \ne n$.*

*Proof.* We can assume without loss of generality that $X$ is a real space, since a complex space can be regarded as a real space with the same norm.

The proof is by induction. We start by choosing an arbitrary $x_1 \in X$ with $\|x_1\| = 1$, so there exists $x_1^* \in X^*$ with $\|x_1^*\| = 1$ and $x_1^*(x_1) = 1$ (Corollary 5.11).

Now suppose we have chosen $x_1^*, \dots, x_n^*$, each with norm 1 and distant more than 1 from all the others, and associated linearly independent functionals $x_1^*, \dots, x_n^*$ of norm 1 with $x_m^*(x_m) = 1$ for $1 \le m \le n$.

Choose $y \in X$ such that $x_m^*(y) < 0$ for $1 \le m \le n$, $x \ne 0$ such that $x_m^*(x) = 0$ for $1 \le m \le n$, and $K \in \mathbb{R}$ such that $\|y\| \le \|y + Kx\|$. Then for any linear combination $\sum_{m=1}^n \alpha_m x_m^*$ where not all the $\alpha_m$ are zero we have

$$\begin{aligned} \left| \sum_{m=1}^n \alpha_m x_m^*(y + Kx) \right| &= \left| \sum_{m=1}^n \alpha_m x_m^*(y) \right| \\ &\le \left\| \sum_{m=1}^n \alpha_m x_m^* \right\| \|y\| \\ &< \left\| \sum_{m=1}^n \alpha_m x_m^* \right\| \|y + Kx\| \end{aligned}$$

(the strict inequality is true in the last line because the $x_m^*$ are linearly independent and the $\alpha_m$ are not all zero, as well as the choice of $K$). Now let $x_{n+1} = (y + Kx)/\|y + Kx\|$ and choose $x_{n+1}^*$ with $\|x_{n+1}^*\| = 1$ and $x_{n+1}^*(x_{n+1}) = 1$ (Corollary 5.11 again). The strict inequality above shows that $x_{n+1}^*$ is not a linear combination of $x_1^*, \dots, x_n^*$, and we have

$$\|x_{n+1} - x_m\| \geq |x_m^*(x_{n+1} - x_m)| = |x_m^*(x_{n+1}) - x_m^*(x_m)|$$

which is greater than $1$ because $x_m^*(x_{n+1}) < 0$ and $x_m^*(x_m) = 1$. □

In case you are worried about why we can always choose vectors $x$ and $y$ with the required properties, it is possible because of the purely algebraic fact that:

**Lemma.** *If $\phi_1, \dots, \phi_n$ and $\phi$ are linear functionals on a vector space $X$, then $\phi$ vanishes on $\bigcap_{m=1}^n \ker(\phi_m)$ if and only if $\phi$ is a linear combination of $\phi_1, \dots, \phi_n$.*

So, in the notation of the proof, if there was no non-zero vector $x$ with $x_m^*(x) = 0$ for $1 \leq m \leq n$, then every $x^* \in X^*$ would vanish on $\bigcap_{m=1}^n \ker(x_m^*)$ and by the lemma would be a linear combination of $x_1^*, \dots, x_n^*$. But then $X^*$ would be finite-dimensional, so $X$ would also be finite-dimensional, contrary to hypothesis.

Similarly, for each $m$ there is a vector $y_m$ such that $x_m^*(y_m) \neq 0$ but each other $x_k^*(y_m) = 0$, for otherwise $x_m^*$ would be a linear combination of the other $x_k^*$, contrary to hypothesis. Without loss of generality $x_m^*(y_m) < 0$, then $y = \sum_{m=1}^n y_m$ has $x_m^*(y) < 0$ for $1 \leq m \leq n$.

*Proof of Lemma.* 'If' is obvious. 'Only if': this is easy for $n = 1$. Assume it is true for some given $n$, and that $\phi$ vanishes on $\bigcap_{m=1}^{n+1} \ker(\phi_m)$. Apply the induction hypothesis to the restrictions of $\phi$ and $\phi_1, \dots, \phi_n$ to $\ker(\phi_{n+1})$ to conclude that, on $\ker(\phi_{n+1})$, $\phi$ is a linear combination of $\phi_1, \dots, \phi_n$, or equivalently that for some $\alpha_1, \dots, \alpha_n$, $\phi - \sum_{m=1}^n \alpha_m \phi_m$ vanishes on $\ker(\phi_{n+1})$. Now we are in the $n = 1$ case again, and can conclude that $\phi - \sum_{m=1}^n \alpha_m \phi_m = \alpha_{n+1}\phi_{n+1}$ for some $\alpha_{n+1} \in \mathbb{R}$. □

## References


1. Joseph Diestel, *Sequences and series in Banach spaces*, Springer-Verlag, 1984.
2. P. R. Halmos, *Naive set theory*, van Nostrand, 1960.
3. Edwin Hewitt and Karl Stromberg, *Real and abstract analysis*, Springer-Verlag, 1969.
4. R. C. James, *A non-reflexive Banach space isometric with its second conjugate space*, Proc. Nat. Acad. Sci. U.S.A. **37** (1951), 174–177.
5. W. Rudin, *Real and complex analysis*, McGraw-Hill, 1966.


DEPARTMENT OF MATHEMATICS, UNIVERSITY OF YORK